ZooKeys 365: 337–354 (2013)
doi: 10.3897/zookeys.365.5356
www.zookeys.org

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Revisiting species delimitation within the genus *Oxystele* using DNA barcoding approach

Herman Van Der Bank[1], Dai Herbert[2], Richard Greenfield[1], Kowiyou Yessoufou[3]

1 *Department of Zoology, African Centre for DNA Barcoding (ACDB), Kingsway Campus, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa* **2** *KwaZulu-Natal Museum, P. Bag 9070, Pietermaritzburg 3200, South Africa, and School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, 3206 South Africa* **3** *Department of Botany and Plant Biotechnology, African Centre for DNA Barcoding (ACDB), Kingsway Campus, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa*

Corresponding author: *Herman Van Der Bank* (hvdbank@uj.ac.za)

Academic editor: *K. Jordaens* | Received 19 April 2013 | Accepted 13 August 2013 | Published 30 December 2013

Citation: Van Der Bank HF, Herbert D, Greenfield R, Yessoufou K (2013) Revisiting species delimitation within the genus *Oxystele* using DNA barcoding approach. In: Nagy ZT, Backeljau T, De Meyer M, Jordaens K (Eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research. ZooKeys 365: 337–354. doi: 10.3897/zookeys.365.5356

Abstract

The genus *Oxystele*, a member of the highly diverse marine gastropod superfamily Trochoidea, is endemic to southern Africa. Members of the genus include some of the most abundant molluscs on southern African shores and are important components of littoral biodiversity in rocky intertidal habitats. Species delimitation within the genus is still controversial, especially regarding the complex *O. impervia* / *O. variegata*. Here, we assessed species boundaries within the genus using DNA barcoding and phylogenetic tree reconstruction. We analysed 56 specimens using the mitochondrial gene COI. Our analysis delimits five molecular operational taxonomic units (MOTUs), and distinguishes *O. impervia* from *O. variegata*. However, we reveal important discrepancies between MOTUs and morphology-based species identification and discuss alternative hypotheses that can account for this. Finally, we indicate the need for future study that includes additional genes, and the combination of both morphology and genetic techniques (e.g. AFLP or microsatellites) to get deeper insight into species delimitation within the genus.

Keywords

Mollusca, Gastropoda, Trochidae, species delimitation, morphology

Copyright *Herman Van Der Bank et al.* This is an open access article distributed under the terms of the Creative Commons Attribution International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Molluscs comprise one of the largest marine phyla, comprising more than 50,000 described species (marine species only), of which less than 10% are currently included in the global database of DNA barcodes (Radulovici et al. 2010). DNA barcoding is a genetic technique designed to standardize and accelerate species identification as an instrument facilitating conservation efforts, ecosystem monitoring, and the identification of phylogeographic and speciation patterns (Radulovici et al. 2010; but see Taylor and Harris 2012 for criticism). It has also proved valuable in population genetics and phylogenetic analyses, identification of prey in gut contents, forensic and seafood safety, invasion biology (Armstrong and Ball 2005, Bucklin et al. 2011) and in revealing cryptic species (Hebert et al. 2004, Puillandre et al. 2009, Lakra et al. 2011). One of the important uses of DNA barcoding is its ability to correctly assign several life-forms including larvae, carcass fragments and damaged specimens to species (Ward et al. 2005, Yang et al. 2012).

Although the mitochondrial cytochrome *c* oxidase subunit I gene (COI), used for barcoding purposes of animals is not efficient for all taxonomic groups (e.g. terrestrial gastropods, Davison et al. 2009; anthozoans, Huang et al. 2008), and pending the integration of the next generation sequencing into the DNA barcoding technique (Taylor and Harris 2012), the barcoding approach has proved valuable in discriminating marine biodiversity (e.g. Sun et al. 2012; see also reviews in Radulovici et al. 2010). *Oxystele* Philippi, 1847, a genus of the highly diverse marine gastropod superfamily Trochoidea (Williams et al. 2010), is endemic to southern Africa. Currently, five species are recognised (Branch et al. 2010), but delimitation within the genus is still debated (Heller and Dempster 1991, Williams et al. 2010), especially due to strong homoplasy in morphological characters traditionally used in identification keys (Hickman 1998).

In this study, our main objective was to infer species boundaries within the genus using DNA barcode. To date, attempts to resolve taxonomic issues within the genus using DNA sequence data were very limited in sample size: only one individual of each of the five recognised *Oxystele* species was generally analysed. For this purpose, we sampled 56 specimens including all five *Oxystele* species from a wide geographic distribution range. We then applied the DNA barcoding approach for taxa delimitation.

Materials and methods

Sample collections

Sampling sites were widely distributed to cover the geographical distribution range of the genus. Species identification was done using the morphological characters given in the key to *Oxystele* species provided by Heller and Dempster (1991). Collection details including GPS coordinates, altitude and photographs of specimens are available online

in the Barcode of Life Data Systems (BOLD; www.boldsystems.org) together with DNA sequences. Voucher specimens (shells) were also collected and deposited at the KwaZulu-Natal Museum (South Africa).

DNA extraction, amplification and sequencing of DNA barcodes

DNA extraction, polymerase chain reactions (PCR) and sequencing of the COI region (animal DNA barcode) were done at the Canadian Centre for DNA Barcoding (CCDB). PCR reactions followed standard CCDB protocols as described by Hajibabaei et al. (2005). This results in 51 COI DNA sequences being generated. We also included in the DNA matrix five COI sequences that we retrieved from BOLD (DQ numbers in Table 1), making the total sequences analysed to a total of 56 COI sequences. Sequence alignment was performed using Multiple Sequence Comparison by Log-Expectation (MUSCLE vs. 3.8.31, Edgar 2004). GenBank accession numbers, BOLD process identification numbers and voucher information are all available online (www.boldsystems.org). These numbers, together with authorities for the species studied are listed in Table 1.

Data analysis

We assessed the "DNA barcode gap" (Meyer and Paulay 2005) in the dataset using two approaches. First, we compared the median of interspecific distances with that of intraspecific distances (genetic distances are calculated between morphospecies). Significance of the differences between both distances was assessed using the non-parametric Wilcoxon ranked sum test. Second, we used Meier et al.'s (2008) approach, that is, we compared the smallest interspecific distance with the largest intraspecific distance. Genetic distances were measured using the Kimura 2-parameter (K2P) model (Kimura 1980). We are aware of the recent literature indicating that the K2P-model might not be the best model for DNA barcoding. However, we used this model here to allow comparison of our results with other DNA barcoding studies where K2P-model is the most frequently used model.

We also tested the discriminatory power of DNA barcoding by evaluating the proportion of correct species identification using the COI region. All sequences were labeled according to the names of the species from which the sequences were generated. The test of discriminatory power works as follows. Each sequence is considered as an unknown while the remaining sequences in the dataset are considered as the DNA barcode database used for identification. If the identification of the query is the same as the pre-considered identification (i.e. the sequence labels), the identification test is scored as "correct", and the overall proportion of correct identification corresponds to the discriminatory power of the region tested, i.e. COI. This test was done applying three approaches: the "best close match" (Meier et al. 2006), the "near neighbour" and the BOLD criteria using respectively the functions bestCloseMatch, threshID,

Table 1. Species, authority, GenBank accession numbers (*DQ*) and BOLD process ID numbers (HVDBM) of specimens studied. Specimens in bold are those for which morphological characters (weathered shell colours and patterns) failed to provide accurate identification; this is revealed in the barcoding test of species delimitation and in phylogenetic tree topology. Sample localities for *O. impervia* and *O. variegata* individuals are indicated: southern Cape[1], Robben Island[2], north-western Cape[3], Namibia[4]

Species (authority):	GenBank and process ID numbers of specimens included in this study	Composition of MOTUs based on the barcoding test of species delimitation
Oxystele sinensis (Gmelin, 1791)	DQ061089, HVDBM056-10, HVDBM083-10, HVDBM084-10, HVDBM085-10, HVDBM086-10, HVDBM087-10, HVDBM409-11, HVDBM410-11, HVDBM411-11, HVDBM412-11, HVDBM437-11	DQ061089, HVDBM056-10, HVDBM083-10, HVDBM084-10, HVDBM085-10, HVDBM086-10, HVDBM087-10, HVDBM409-11, HVDBM410-11, HVDBM411-11, HVDBM412-11, HVDBM437-11
Oxystele tabularis (Krauss, 1848)	DQ061090, HVDBM289-11, HVDBM338-11, HVDBM339-11	DQ061090, HVDBM289-11, HVDBM338-11, HVDBM339-11
Oxystele tigrina (Anton, 1838)	DQ061091, HVDBM005-10, HVDBM006-10, HVDBM013-10, HVDBM055-10, HVDBM394-11, HVDBM506-11, HVDBM507-11, HVDBM508-11, HVDBM509-11, HVDBM510-11	DQ061091, HVDBM005-10, HVDBM006-10, HVDBM013-10, HVDBM055-10, HVDBM394-11, HVDBM506-11, HVDBM507-11, HVDBM508-11, HVDBM509-11, HVDBM510-11
Oxystele variegata (Anton, 1838)	DQ061092[1], HVDBM058-10[1], HVDBM059-10[1], HVDBM070-10[1], HVDBM072-10[1], HVDBM183-10[3], HVDBM184-10[3], HVDBM185-10[3], HVDBM208-10[4], HVDBM209-10[4], HVDBM389-11[3], HVDBM393-11[1], HVDBM395-11[1], HVDBM456-11[4], HVDBM457-11[4], HVDBM511-11[2], HVDBM512-11[2], HVDBM513-11[2], HVDBM514-11[2], HVDBM515-11[2]	HVDBM072-10[1], HVDBM183-10[3], HVDBM184-10[3], HVDBM185-10[3], HVDBM208-10[4], HVDBM209-10[4], HVDBM389-11[3], HVDBM393-11[1], HVDBM395-11[1], HVDBM456-11[4], HVDBM457-11[4], HVDBM511-11[2], HVDBM512-11[2], HVDBM513-11[2], HVDBM514-11[2], HVDBM515-11[2], **HVDBM028-10[1]**
Oxystele impervia (Menke, 1843)	DQ061093[1], HVDBM022-10[1], HVDBM027-10[1], HVDBM028-10[1], HVDBM057-10[1], HVDBM071-10[1], HVDBM178-10[3], HVDBM179-10[3], HVDBM180-10[3]	**DQ061093**[1], HVDBM022-10[1], HVDBM027-10[1], HVDBM057-10[1], HVDBM071-10[1], HVDBM178-10[3], HVDBM179-10[3], HVDBM180-10[3], **DQ061092**[1], **HVDBM058-10[1]**, **HVDBM059-10[1]**, **HVDBM070-10[1]**

and nearNeighbour implemented in the program Spider 1.1-1 (Brown et al. 2012). Prior to the test, we determined the optimised genetic distance suitable as threshold for taxon identification. For this purpose, we used the function localMinima also implemented in Spider (Brown et al. 2012).

The function bestCloseMatch conducts the "best close match" analysis of Meier et al. (2006), searching for the closest individual in the dataset. If the closest individual is within a given threshold, the outcome is scored as "correct". If it is further than the given threshold, the result is "no ID" (no identification). If more than one species are tied for closest match, the outcome of the test is "ambiguous" identification. When all matches within the threshold are different species to the query, the result is scored as "incorrect".

The function threshID conducts a threshold-based analysis based on a threshold genetic distance of 1% as conducted by the "Identify Specimen" tool provided by the BOLD system (http://www.boldsystems.org/views/idrequest.php). It is more inclusive than bestCloseMatch, in that it considers all sequences within the threshold of 1%. There also four possible outcomes for threshID tests, that is, "correct", "incorrect", "ambiguous", and "no id" similar to the outcomes of the bestCloseMatch function.

The nearNeighbour function finds the closest individual and returns the score "true" (equivalent to "correct") if their names are the same, but if the names are different, the outcome is scored as "false" (equivalent to "incorrect").

Further, we performed a barcoding test of taxon delimitation. In reality, this test groups specimens into "molecular operational taxonomic units" (MOTUs; Jones et al. 2011), which are generally regarded as proxy for morpho-species (Stahlhut et al. 2013). MOTUs are defined as groups of specimens that are within the genetic threshold used for taxon delimitation. If all specimens of the same morpho-species are clustered in a single MOTU, this means that MOTUs are congruent with morpho-species, thus increasing the taxonomic value of DNA barcoding. The delimitation of MOTUs was conducted using the function tclust in the R package Spider v1.1-1. If two specimens are more distant than the threshold from each other, but both are within the threshold of a third, the function tclust identified all three individuals as a single MOTU. We also identified the composition of each MOTU using the function lapply also implemented in Spider.

Finally we complemented the test of MOTU delimitation with a phylogenetic analysis of COI sequences. We reconstructed a phylogenetic tree using Bayesian and maximum parsimony methods. The Bayesian tree was reconstructed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). The best-fit model of DNA sequence evolution was chosen using jModelTest v0.1.1 (Posada 2008) under the Akaike information criterion (Posada and Buckley 2004). The TrN + I model was selected and used to generate the Bayesian tree. Analysis was run for nine million generations with sampling one tree every 100 generations. Two independent Bayesian analyses with four differentially heated chains were performed simultaneously. The results were visualised and checked using MEGA, and 25,000 trees were discarded as burn-in to ensure that the analysis had stabilised. Node support was assessed using posterior probability (PP) as follows: PP > 0.95: high support and PP < 0.95: no support (Alfaro and Holder 2006).

Maximum parsimony (MP) was implemented to analyse the data using PAUP* v4.10b10 (Swofford 2002). Tree searches were done using heuristic searches with 1000 random sequence additions but keeping only 10 trees. Tree bisection-reconnection was performed with all character transformations treated as equally likely i.e. Fitch parsimony (Fitch 1971). MP searches and bootstrap resampling (Felsenstein 1985) were done using PAUP* v4.10b10 (Swofford 2002).

Jujubinus exasperatus (Pennant, 1777) was used as outgroup based on Williams et al. (2010). Node support was assessed using bootstrap (BP) values: BP > 70% for strong support (Murphy et al. 2001, Wilcox et al. 2002).

Results

Our dataset includes 56 specimens: nine specimens of *O. impervia,* 12 of *O. sinensis,* four of *O. tabularis*, 11 of *O. tigrina*, and 20 specimens of *O. variegata* (Table 1). The aligned COI matrix was 654 base pairs in length, including A: 24.2%; C: 21.1%; G: 18.3% and T: 36.4%.

Interspecific distances range from 0 to 0.18 (median = 0.15) and are generally larger than intraspecific distances (range: 0-0.09; median = 0.004; Wilcoxon test, $p < 0.001$; Figure 1A). This indicates that there is a barcode gap in the dataset. Even when we compared the lowest interspecific versus the furthest intraspecific distance, we also found that barcode gap exists within the COI sequences (grey lines in Figure 1B).

We determined the optimised threshold genetic distance (d) with which we tested the discriminatory power of COI sequences and delimited MOTUs. We found d = 0.047 (Figure 2). Testing the efficacy of DNA barcoding based on this threshold,

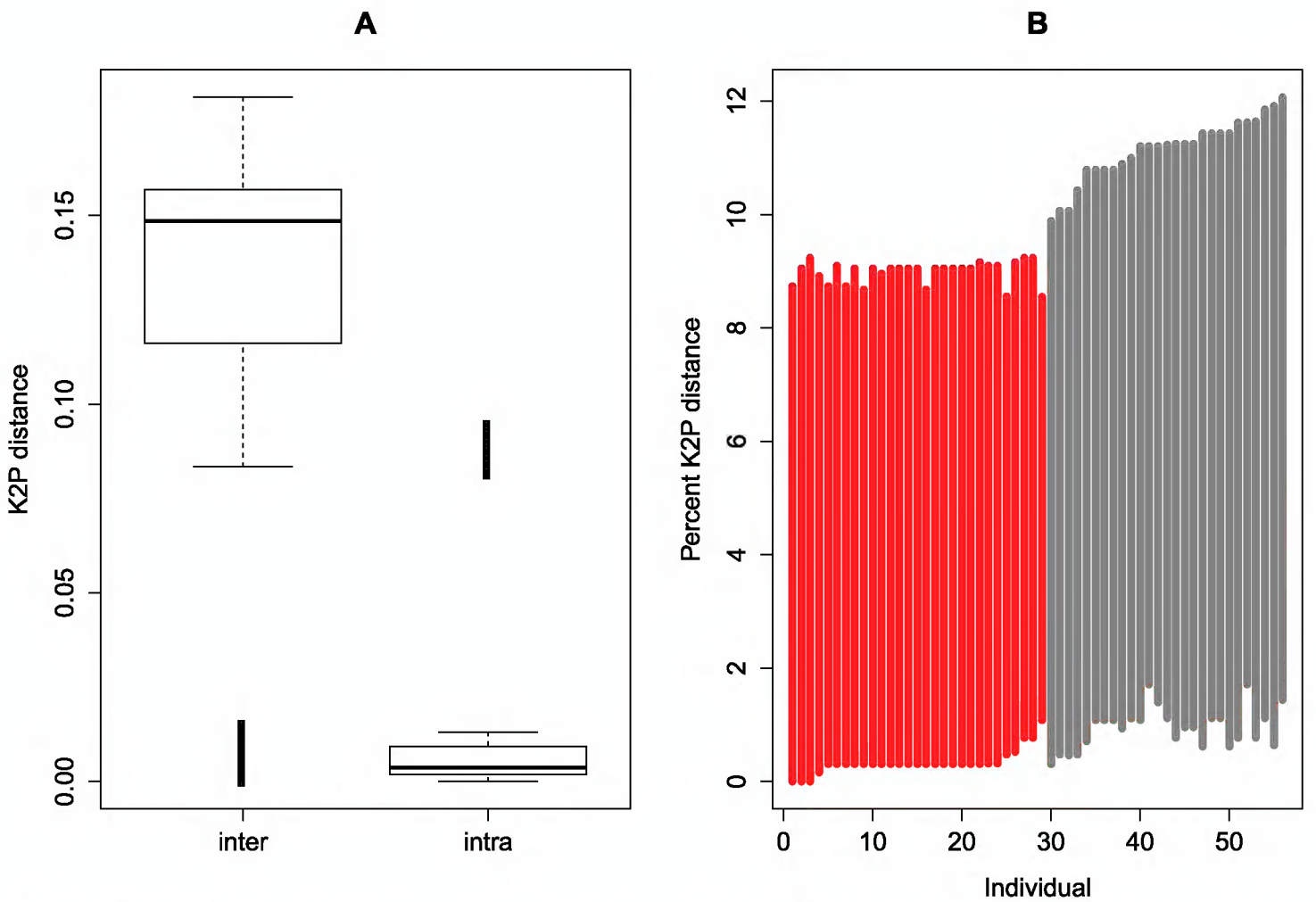

Figure 1. Evaluation of barcode gap in the dataset. **A** Boxplot of the interspecific (inter) and intraspecific genetic (intra) distances, indicating the existence of a barcode gap i.e. intraspecific distance is longer than intraspecific distance. The bottom and top of the boxes show the first and third quartiles respectively, the median is indicated by the horizontal line, the range of the data by the vertical dashed line and outliers (points outside 1.5 times the interquartile range) by Bold vertical lines **B** Lineplot of the barcode gap for the 56 *Oxsystele* specimens. For each specimen in the dataset, the grey lines indicate where the smallest interspecific distance (top of line value) is longer than the longest intraspecific distance (bottom of line value), therefore indicating existence of barcode gap; the red lines show where this pattern is reversed, and the closest non-conspecific is closer to the query than its nearest conspecific, i.e., the situation where there is no barcoding gap.

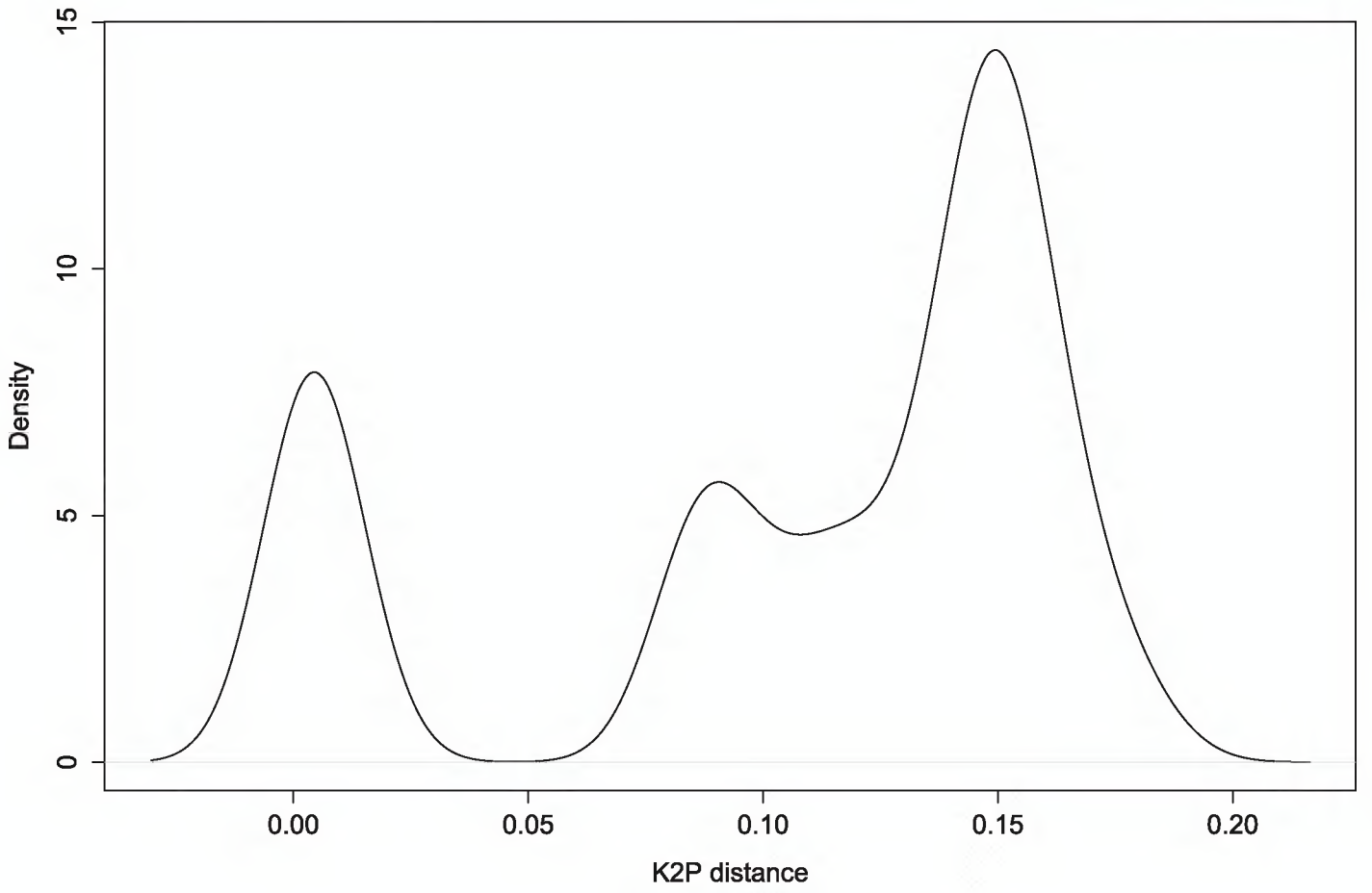

Figure 2. Determination of the threshold genetic distance for species identification. The density plot indicates transition between intra- and interspecific distances; the genetic distance corresponding to this transition (dip in the density graph, here approximately 0.05) indicates the suitable threshold to the dataset. This method does not require prior knowledge of species identity to get an indication of potential threshold values.

Table 2. Tests of barcoding identification accuracy with numbers (n) and percentages (%) of each score.

Methods	**Near neighbour**		**Best Close match**				**BOLD criteria**			
Scores	False	True	Ambiguous	Correct	Incorrect	No ID	Ambiguous	Correct	Incorrect	No ID
n (%)	7 (12.5%)	49 (87.5%)	3 (5.36%)	49 (87.5%)	4 (7.14%)	0 (0%)	27 (48.21%)	27 (48.21%)	1 (1.79%)	1 (1.79%)

we found that COI sequences performed very well in assigning DNA sequences to the correct species (Table 2). For instance, under both near neighbour and best close match methods, 87.5% of the COI sequences were correctly identified (49 specimens out of 56). However, the best close match method indicates 5.36% of ambiguity (three specimens), i.e. both correct and incorrect species are within the given threshold; and 7.14% of incorrect identification (four specimens). Also, for 12.5% of sequences (seven specimens) the near neighbour method results in "incorrect". Using the BOLD method (threshold = 1%), we obtained poor barcoding performance, that is, we have as many correct as ambiguous results (48.21% respectively; i.e. 27 specimens). The BOLD method also indicates one "incorrect" and one "no id" (Table 2).

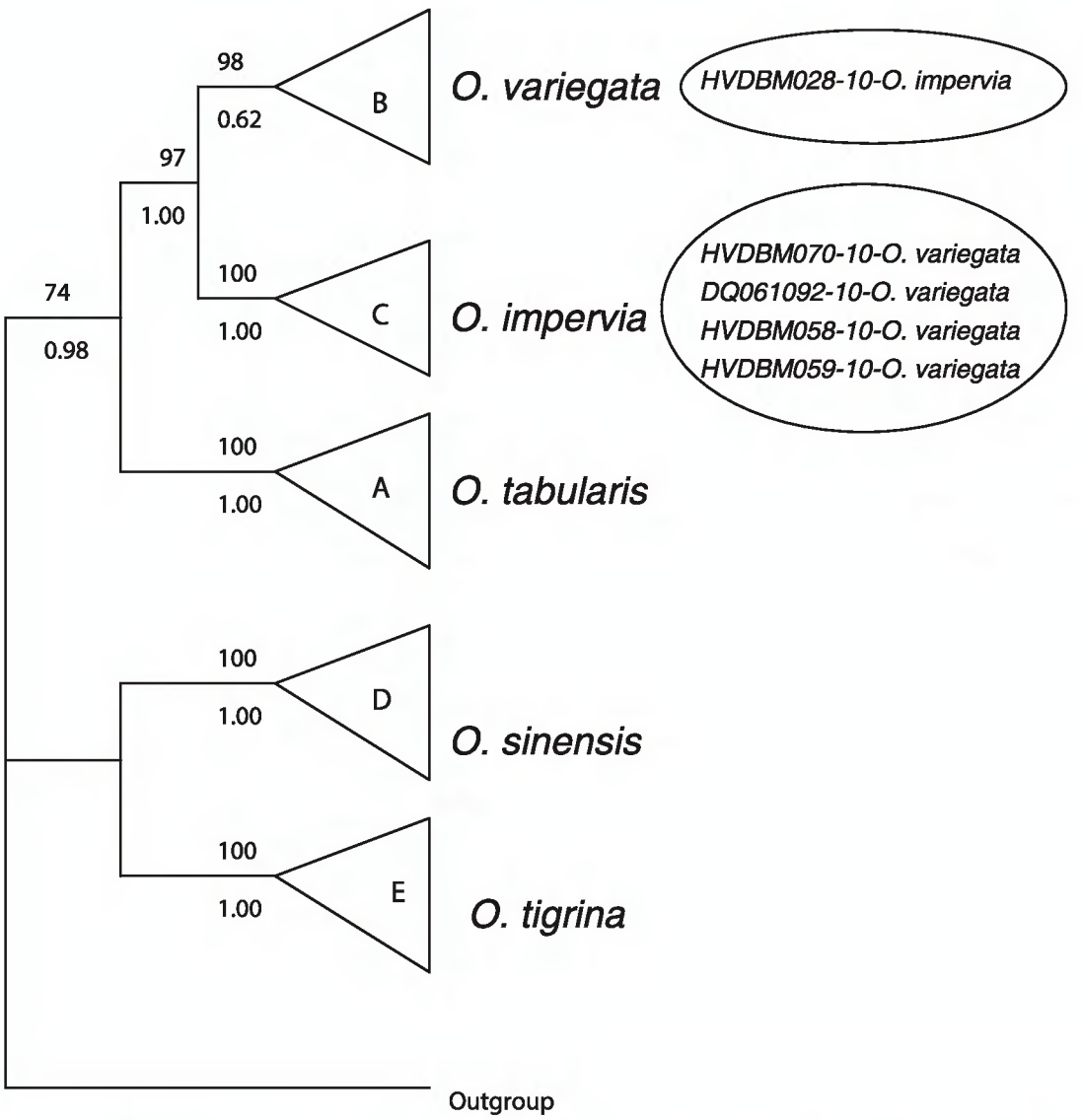

Figure 3. Summary of both Bayesian and parsimonious trees. Values above branches indicate bootstrap supports; values under branches indicate posterior probability. All distinguished species are indicated at the tip of the tree. Branches without values indicate non-supported nodes; the small circle indicates a specimen of *O. impervia* (HVDBM028-10) that was misidentified based on morphology; large circle indicates four specimens morphologically indistinguishable from *O. variegata* (HVDBM070-10; DQ061092; HVDBM058-10; HVDBM059-10), but that are, based on both barcoding analysis of species delimitation (see Table 1) and phylogenetic tree analysis identified as *O. impervia* (see also Appendices 1 and 2).

Further, all the 56 specimens included in this study were grouped into five MOTUs based on our threshold (Table 1). Using tree-based analysis, we also found five strongly supported groupings (PP = 1.00; BP = 100%), identified as A–E (Figure 3), except that the grouping B corresponding to *O. variegata* is only well supported in the MP analysis (BP = 98%). The composition of these five groupings matches that of MOTUs and comprises *O. tabularis* (A), *O. variegata* (B), *O. impervia* (C), *O. sinensis* (D), and *O. tigrina* (E) (Figures 3, Appendix 1 and 2).

Discussion

The concept of DNA barcoding was first proposed as a technique to accelerate species identification within micro-organisms (Nanney 1982). However, it has now been

generalised as a potential method that can help characterise and discover new species in broader taxonomic groups (Hebert et al. 2004, Van der Bank et al. 2012). In the animal kingdom, the COI region has proved valuable as a DNA barcode for many taxonomic groups, but it can also be problematic for others (Moritz 2004, Ebach and Holdrege 2005, Schindel and Miller 2005, Köhler 2007, Huang et al. 2008).

We first tested COI's potential as a good barcode for the genus *Oxystele*. A good barcode candidate is expected to exhibit a barcode gap (Meyer and Paulay 2005), i.e. higher genetic variation between than within species (Hebert et al. 2003). Various options are currently available to evaluate the barcode gap. We used two approaches. We compared the median of interspecific versus intraspecific distances. We found that interspecific distance is significantly greater than intraspecific distance, suggesting that there is a barcode gap in COI data. We also applied the approach of Meier et al. (2008); i.e. compared the smallest interspecific versus the greatest intraspecific distances), rather than comparing just the median distances. This approach also reveals existence of a barcode gap, thus confirming COI as a potential DNA region for taxon identification within *Oxystele*. This DNA region has also proved successful for barcoding identification in other mollusc taxonomic groups (Davison et al. 2009, Köhler and Glaubrecht 2009, Feng et al. 2011a,b, Sun et al. 2012; but see Sauer and Hausdorf 2012 for limitation of single-locus DNA sequences).

In addition, we found that COI has a strong discriminatory power (85%) within the genus *Oxystele* especially using the best close match and near neighbour methods. This gives support to the efficacy of COI for identification purposes within the genus. However, the application of BOLD identification criteria yields a poor identification success i.e. < 50% and similar proportion of ambiguity (Table 2). The poor performance of COI using BOLD criteria should not be seen as a result of barcoding inefficiency, but should rather be linked to the untested 1% threshold used in BOLD identification (see Meyer and Paulay 2005).

Our analysis of barcoding-based taxon delimitation results in five MOTUs, of which three correspond to morphology-delimited species: *O. sinensis, O. tabularis* and *O. tigrina* (Table 1). These results are also supported by phylogeny-based analysis of species delimitation. However, four specimens identified morphologically as *O. variegata* are included by the barcoding taxon delimitation test within the MOTU of *O. impervia*. Similarly, one specimen identified morphologically as *O. impervia* is grouped within the MOTU of *O. variegata* (Figure 3). These mismatches between morpho-species identification and barcoding-based taxon delimitation (MOTUs) reflect the controversy surrounding species boundaries and/or the identification key (e.g. Heller and Dempster's (1991) key) currently used to distinguish the *impervia/variegata* complex.

Why the mismatch between MOTU and morpho-species? Potential explanations include unsuitable morphology-based taxon delimitation, species paraphyly (– including but not restricted to ancestral polymorphism), and on-going gene flow (i.e., the two taxa are not distinct species or they hybridize; see Funk and Omland 2003). Specifically, Funk and Omland (2003) demonstrated that about 25% of animal species are para- or even polyphyletic, suggesting that the non-monophyly of *O. variegata* and *O.*

impervia in the examined gene tree is not necessarily an argument against their species status. This provides further evidence of the limitations of DNA barcoding in general. It is also possible that the rate of speciation events is slower or greater than that of morphological differentiation; e.g. rapid morphological changes can occur with little or no evolutionary changes (Adams et al. 2002); and this could be driven for example by habitat specialisation (Collar et al. 2010).

In our attempt to resolve the taxonomic uncertainty, we also used the phylogenetic tree reconstruction. The results are similar to those of MOTUs, that is, one specimen morphologically identified as *O. impervia,* grouped on the phylogeny with *O. variegata* (grouping B, Figure 3, Appendix 1 and 2), but this grouping B has strong support only in MP analysis.

The controversy regarding the complex has been reported in previous studies (Heller and Dempster 1991, Williams et al. 2010), likely reflecting the limitations in morphological characters (Hickman 1998) on which the current identification key is based. Heller and Dempster (1991) reported that *O. impervia* and *O. variegata* should be considered as two different species based on shell colour, radula cusp indentation, ecological (*O. impervia* occurs higher up the shore than *O. variegata)*, and fixed allozyme differences at one enzyme-coding locus (out of 22). However, the overlaps in ecological zones and interspecific overlap of up to 66% in radula cusp indentation (Heller and Dempster 1991) indicate that these criteria (ecology and radula indentation) might be unreliable for taxon identification.

In addition, Heller and Dempster (1991) described 24 different photos of shell colours and patterns of typical *O. impervia* and *O. variegata* (12 photos for each species), but the differentiation they proposed is still unclear and could lead to multiple interpretations as indicated in the words such as "very infrequently", "off-white", or "greenish-grey" and "almost never" that they used to distinguish between both species. Also, overlaps in colours and weathered shells make Heller and Dempster's (1991) keys unreliable to identify some individuals (e.g. see Figure 4). Specimens of both *O. impervia* and *O. variegata* are commonly weathered to some extent, resulting in shell colour being indistinct or scarcely discernible. Some specimens (e.g. as shown in Figure 4) can only be tentatively identified because they exhibit unusual colour patterns, not clearly consistent with published photos in Heller and Dempster (1991).

Williams et al. (2010) however suggested that *O. impervia* and *O. variegata* should be regarded as one species based on analysis from a single individual from each species. DH inspected the morphology of the samples (available on MorphoBank) used in the study by Williams et al. (2010) and confirmed that the shell of specimen DQ061092-10 is very typical of that of *O. variegata*, but that DQ061093-10 has a more intermediate form with a finer colour pattern. He concluded that the latter is not obviously referable to any one of *O. impervia* and/or *O. variegata,* more than to the other. In this study, the fact that both specimens come out not only on the phylogeny in the grouping of *O. impervia* (grouping C on the phylogeny; with strong support from PP and BP; Figures 3, Appendix 1 and 2), but also in the MOTU delimitation (Table 1), is surprising (particularly DQ061092-10, which is morphologically typical of *O. variegata*).

Figure 4. Patterns of shell colour within the genus *Oxystele*. **A–C** *Oxystele variegata* from Namibia, 5 km north of Swakopmund, diameter 22.2 mm (NMSA E6038) **D–F** *Oxystele impervia* from the Western Cape, Groen Rivier, diameter 22.3 mm (NMSA E7353) **G–I** *Oxystele* sp. from the Eastern Cape, Tsitsikamma National Park, diameter 16.5 mm (HVDBM058-10, NMSA W7371); the colour pattern of these specimens suggests *O. variegata*, but these specimens group within the unit of *O. impervia* **J–L** *Oxystele* sp. from the Northern Cape, Noup, diameter 18.0 mm (HVDBM185-10, NMSA W7608); the colour pattern suggests *O. impervia*, but they group with *O. variegata* (see Figures 4 and Appendix 2 for the phylogenetic groupings of these specimens and node supports; these groupings contradict their morphological identification).

One of six polymorphic loci (glycyl-leucine peptidase or peptidase A; Van der Bank 2002) indicated fixed allele differences between *O. impervia* and *O. variegata*, and this was the most convincing characteristic to differentiate between both species (Heller and Dempster 1991). Williams et al. (2010) argue that differences in allele frequency could result from selection pressures (e.g. peptidase in *Mytilus*; Hilbish 1985). They further indicate that differences in habitat preferences, as reported for the *impervia/variegata* complex, could subject them to variation in salinity or temperature, which could lead to variation not only in diets but also in allozymes and morphology.

Indeed morphological differentiation between both species can be difficult. Some of the shell colours and patterns are similar, and radula morphology could be altered

as a result of differences in diet, age and other factors. For example, Padilla (1998) demonstrated that two species of Gastropoda "produce differently shaped teeth when fed different foods, displaying intraspecific variability as extreme as would usually be considered to define different species". Such variation in morphological characters has also been reported to be misleading in other groups such as spiders where the description of almost 50% of the known species was mistakenly based on the same species (Coddington and Levi 1991). Indeed molluscs are well-known to exhibit considerable intraspecific variation in shell morphology (Colgan et al. 2007; Figure 4), and high adaptive capacity to various environmental conditions, leading to striking ecological, morphological and behavioural disparity among specimens within the same species (Ponder et al. 2008).

In this study, most of the specimens that group within unexpected MOTUs were collected from different localities, suggesting possible shell colour variation due to variation in environmental conditions. For example, specimens of *O. variegata* from Namibia and Robben Island clustered on the phylogeny, but those from north-western and southern Africa (Cape) did not. The Cape is renowned for its bad weather as indicated in its common name of "The Cape of Storms", resulting in weathering of individuals (i.e. see "Ships in trouble in Cape waters"; http://www.e-gnu.com/shipwreck_update.html).

Conclusion

The split we found on the phylogeny and species delimitation analyses between *O. impervia* and *O. variegata* does not correspond with the nominal, morphologically-based identifications, indicating the need for the combination of morphological features and genetic data for further analysis. It is also possible that the COI gene alone is insufficient to discriminate species within the genus. We therefore suggest that future analysis should use a multi-gene approach. However, Donald et al. (2005) have studied three genes including two mitochondrial (16S + COI) and one nuclear (actin), and Williams et al. (2010) used one nuclear and three mitochondrial genes; but neither study was successful in teasing apart both species. We would therefore suggest that additional techniques such as AFLP or microsatellites should be applied in an attempt to reveal the status of *O. impervia* and *O. variegata*. Nevertheless, our analyses using barcoding confirm the existence of five MOTUs (probably suggestive of five species), with *O. variegata* being a distinct species from *O. impervia*.

Acknowledgements

We would like to thank the Government of Canada through Genome Canada and the Ontario Genomics Institute (2008-OGI-ICI-03) for the DNA sequencing. The research was supported by the ACDB and partially by the Toyota Enviro Outreach program

2010. The Tsitsikamma National Parks Board gave permission for sample collection and the Kwazulu-Natal Museum processed the voucher specimens. We thank Stephanus Voges and Bronwen Curry (Ministry of Fisheries and Marine Resources, Namibia), Estelle Esterhuizen (Robben Island Nature Conservation) and Gerhard Groenewald (Klipbokkop Nature Reserve) for assistance with sample collections. This work is based on research supported in part by the National Research Foundation of South Africa.

References

Adams PA, Pandey N, Rezzi S, Casanova J (2002) Geographic variation in the Random Amplified Polymorphic DNAs (RAPDs) of *Juniperus phoenicea*, *J. p.* var. *canariensis*, *J. p.* subsp. *eumediterranea*, and *J. p.* var. *turbinata*. Biochemical Systematic Ecology 30: 223–229. doi: 10.1016/S0305-1978(01)00083-7

Alfaro ME, Holder MT (2006) The posterior and the prior in Bayesian Phylogenetics. Annual Review of Ecology Evolution and Systematics 37: 19–42. doi: 10.1146/annurev.ecolsys.37.091305.110021

Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Philosophical Transactions of the Royal Society B 360: 1813–1823. doi: 10.1098/rstb.2005.1713

Branch GM, Griffiths CL, Branch ML, Beckley LE (2010) Two oceans: a guide to the marine life of southern Africa. Struik Nature, Cape Town.

Brown SDJ, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources. http://cran.r-project.org; http://spider.r-forge.r-project.org/, doi: 10.1111/j.1755-0998.2011.03108.x

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. Annual Review of Marine Science 3: 471–508. doi: 10.1146/annurev-marine-120308-080950

Coddington JA, Levi HW (1991) Systematics and evolution of spiders (Araneae). Annual Reviews of Ecology and Systematics 22: 565–592. doi: 10.1146/annurev.es.22.110191.003025

Colgan DJ, Ponder WF, Beacham E, Macaranas J (2007) Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). Molecular Phylogenetics and Evolution 42: 717–737. doi: 10.1016/j.ympev.2006.10.009

Collar DC, Schulte JA, O'meara BC, Losos JB (2010) Habitat use affects morphological diversification in dragon lizards. Journal of Evolutionary Biology 23: 1033–1049. doi: 10.1111/j.1420-9101.2010.01971.x

Davison A, Blackie RLE, Scothern GP (2009) DNA barcoding of stylommatophoran land snails: a test of existing sequences. Molecular Ecology Resources 9: 1092–1101. doi: 10.1111/j.1755-0998.2009.02559.x

Donald KM, Kennedy M, Spencer HG (2005) The phylogeny and taxonomy of austral monodontine topshells (Mollusca: Gastropoda: Trochidae), inferred from DNA sequences. Molecular Phylogenetics and Evolution 37: 474–483. doi: 10.1016/j.ympev.2005.04.011

Ebach MC, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. Nature 434: 697. doi: 10.1038/43469

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797. doi: 10.1093/nar/gkh340

Feng Y, Li Q, Kong L, Zheng X (2011a) COI-based DNA barcoding of Arcoida species (Bivalvia: Pteriomorphia) along the coast of China. Molecular Ecology Resources 11: 435–441.

Feng Y, Li Q, Kong L, Zheng X (2011b) DNA barcoding and phylogenetic analysis of Pectinidae (Mollusca: Bivalvia) based on mitochondrial COI and 16S rRNA genes. Molecular Biology Report 38: 291–299. doi: 10.1007/s11033-010-0107-1

Felsenstein J (1985) Confidence levels on phylogenies: an approach using the bootstrap. Evolution 39: 783–791. DOI: 10.2307/2408678

Fitch WM (1971) Towards defining the course of evolution: minimum change for a specific tree topology. Systematic Zoology 20: 406–416. doi: 10.2307/2412116

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology Evolution and Systematics 34: 397–423. doi: 10.1146/annurev.ecolsys.34.011802.132421

Hajibabaei M, De Waard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM, Hebert PDN (2005) Critical factors for assembling a high volume of DNA barcodes. Philosophical Transactions of the Royal Society B 360: 1959–1967. doi: 10.1098/rstb.2005.1727

Hebert PDN, Ratnasingham S, de Waard JR (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London B (Supplement) 270: S96–S99. doi: 10.1098/rsbl.2003.0025

Hebert PDN, Penton EH, Burns J, Janzen DJ, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly, *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the USA 101: 14812–14817. doi: 10.1073/pnas.0406166101

Heller J, Dempster Y (1991) Detection of two coexisting species of *Oxystele* (Gastropoda, Trochidae) by morphological and electrophoretic analysis. Journal of Zoology 223: 395–418. doi: 10.1111/j.1469-7998.1991.tb04773.x

Hickman CJ (1998) A field guide to sea stars and other echinoderms of Galápagos. Sugar Spring Press, Lexington, VA, USA, 83 pp.

Hilbish TJ (1985) Demographic and temporal structure of an allele frequency cline in the mussel *Mytilus edulis*. Marine Biology 86: 163–171. doi: 10.1007/BF00399023

Huang DW, Meier R, Todd PA, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. Journal of Molecular Evolution 66: 167–174. doi: 10.1007/s00239-008-9069-5

Jones M, Ghoorah A, Blaxter M (2011) jMOTU and Taxonerator: Turning DNA barcode sequences into annotated operational taxonomic units. PLoS ONE 6: e19259. doi: 10.1371/journal.pone.0019259

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111–120. doi: 10.1007/BF01731581

Köhler F (2007) From DNA taxonomy to barcoding - how a vague idea evolved into a biosystematic tool. Mitteilungen aus dem Museum für Naturkunde in Berlin Zoologische Reihe 83: 44–51. doi: 10.1002/mmnz.200600025

Köhler F, Glaubrecht M (2009) Uncovering an overlooked radiation: molecular phylogeny and biogeography of Madagascar's endemic river snails (Caenogastropoda: Pachychilidae: *Madagasikara* gen. nov.). Biological Journal of the Linnean Society 99: 867–894. doi: 10.1111/j.1095-8312.2009.01390.x

Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, Gopalakrishnan A, Ward RD, Hebert P (2011) DNA barcoding Indian marine fishes. Molecular Ecology Resources 11: 60–71. doi: 10.1111/j.1755-0998.2010.02894.x

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55: 715–728. doi: 10.1080/10635150600969864

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Systematic Biology 57: 809–813. doi: 10.1080/10635150802406343

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3: 2229–2238. doi: 10.1371/journal.pbio.0030422

Moritz C (2004) DNA barcoding: promise and pitfalls. PLoS Biology 2: e354. doi: 10.1371/journal.pbio.0020354

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E et al. (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294: 2348–2351. doi: 10.1126/science.1067179

Nanney DL (1982) Genes and phenes in *Tetrahymena*. BioScience 32: 783–788. doi: 10.2307/1308971

Padilla DK (1998) Inducible phenotypic plasticity of the radula in *Lacuna* (Gastropoda: Littorinidae). The Veliger 41: 201–204.

Ponder WF, Colgan DJ, Healy JM, Hützel A, Simone LRL, Strong EE (2008) Caenogastropoda. In: Ponder WF, Lindberg DR (Eds) Phylogeny and Evolution of the Mollusca. University of California Press, Berkeley, 331–383. doi: 10.1525/california/9780520250925.003.0013

Posada D (2008) jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256. doi: 10.1093/molbev/msn083

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53: 793–808. doi: 10.1080/10635150490522304

Puillandre N, Cruaud C, Kantor YI (2009) Cryptic species in *Gemmuloborsonia* (Gastropoda: Conoidea). Journal of Molluscan Studies 76: 11–23. doi: 10.1093/mollus/eyp042

Radulovici AE, Archambault P, Dufresne F (2010) DNA barcodes for marine biodiversity: moving fast forward? Diversity 2: 450–472. doi: 10.3390/d2040450

Ronquist F, Huelsenbeck JP (2003) MrBayes 3.1.2: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574. doi: 10.1093/bioinformatics/btg180

Sauer J, Hausdorf B (2012) A comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy. Cladistics 28: 300–316. doi: 10.1111/j.1096-0031.2011.00382.x

Schindel DE, Miller SE (2005) DNA barcoding, a useful tool for taxonomists. Nature 435: 17. doi: 10.1038/43501

Stahlhut JK, Fernández-Triana J, Adamowicz SJ, Buck M, Goulet H, Hebert PDN, Huber JT, Merilo MT, Sheffield CS, Woodcock T, Smith MA (2013) DNA barcoding reveals diversity of Hymenoptera and the dominance of parasitoids in a sub-arctic environment. BMC Ecology 13: 2. doi: 10.1186/1472-6785-13-2

Sun Y, Li Q, Kong L, Zheng X (2012) DNA barcoding of Caenogastropoda along coast of China based on the COI gene. Molecular Ecology Resources 12: 209–218. doi: 10.1111/j.1755-0998.2011.03085.x

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.10. Sinauer, Sunderland, Massachusetts.

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Molecular Ecology Resources 12: 377–388. doi: 10.1111/j.1755-0998.2012.03119.x

Van der Bank FH (2002) A review of gene nomenclature for enzyme-coding loci generally used in allozyme studies. Trends in Comparative Biochemistry and Physiology 9: 197–203. doi: 10.3750/AIP2012.42.4.04

Van der Bank HF, Greenfield R, Daru BH, Yessoufou K (2012) DNA barcoding reveals micro-evolutionary changes and river system-level phylogeographic resolution of African silver catfish, *Schilbe intermedius* (Actinopterygii: Siluriformes: Schilbeidae) from seven populations across different African river systems. Acta Ichthyologica Et Piscatoria 42: 307–320.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B 360: 1847–1857. doi: 10.1098/rstb.2005.1716

Wilcox T, Zwick D, Heath T, Hillis D (2002) Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Molecular Phylogenetics and Evolution 25: 361–371. doi: 10.1016/S1055-7903(02)00244-0

Williams ST, Donald KM, Spencer HG, Nakano T (2010) Molecular systematics of the marine gastropod families Trochidae and Calliostomatidae (Mollusca: Superfamily Trochoidea). Molecular Phylogenetics and Evolution 54: 783–809. doi: 10.1016/j.ympev.2009.11.008

Yang JB, Wang YP, Moller M, Gao LM, Wu D (2012) Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). Molecular Ecology Resources 12: 267–275. doi: 10.1111/j.1755-0998.2011.03095.x

Appendix I

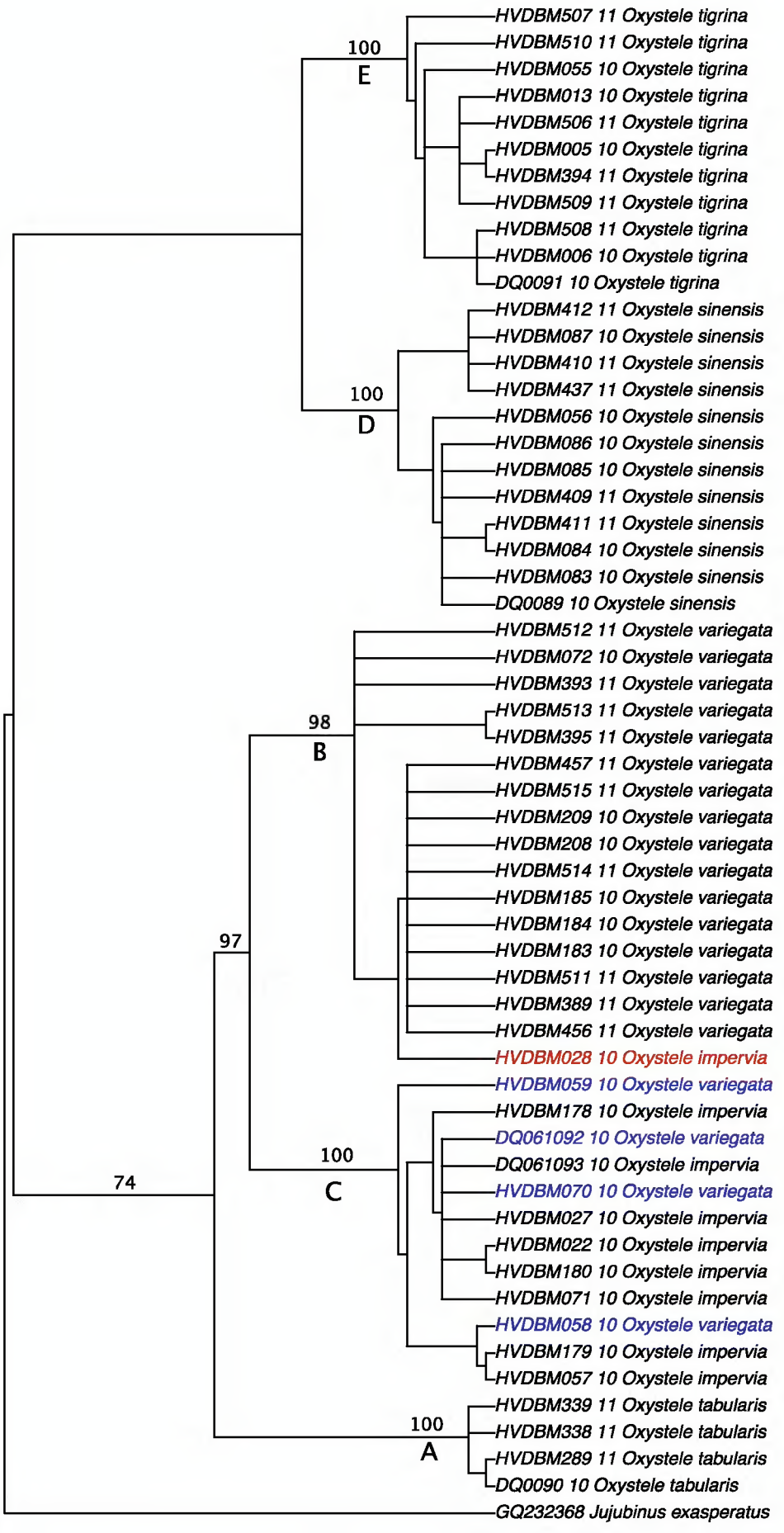

Figure S1. The only parsimonious tree obtained from the maximum parsimony (MP) analysis. Topology of species groupings is similar to that of the Bayesian tree (see Figure 3). Node supports are reported on the branches; the first value is bootstrap support from MP analysis; the second value in bracket indicates the posterior probability obtained from Bayesian analysis; only moderate to high node support values are indicated; *Jujubinus exasperatus* is used as outgroup; A-E indicates different possible species-units in the dataset: A (*O. tabularis*), B (*O. variegata*), C (*O. impervia*), D (*O. sinensis*), E (*O. tigrina*), as in Figure 3.

Appendix 2

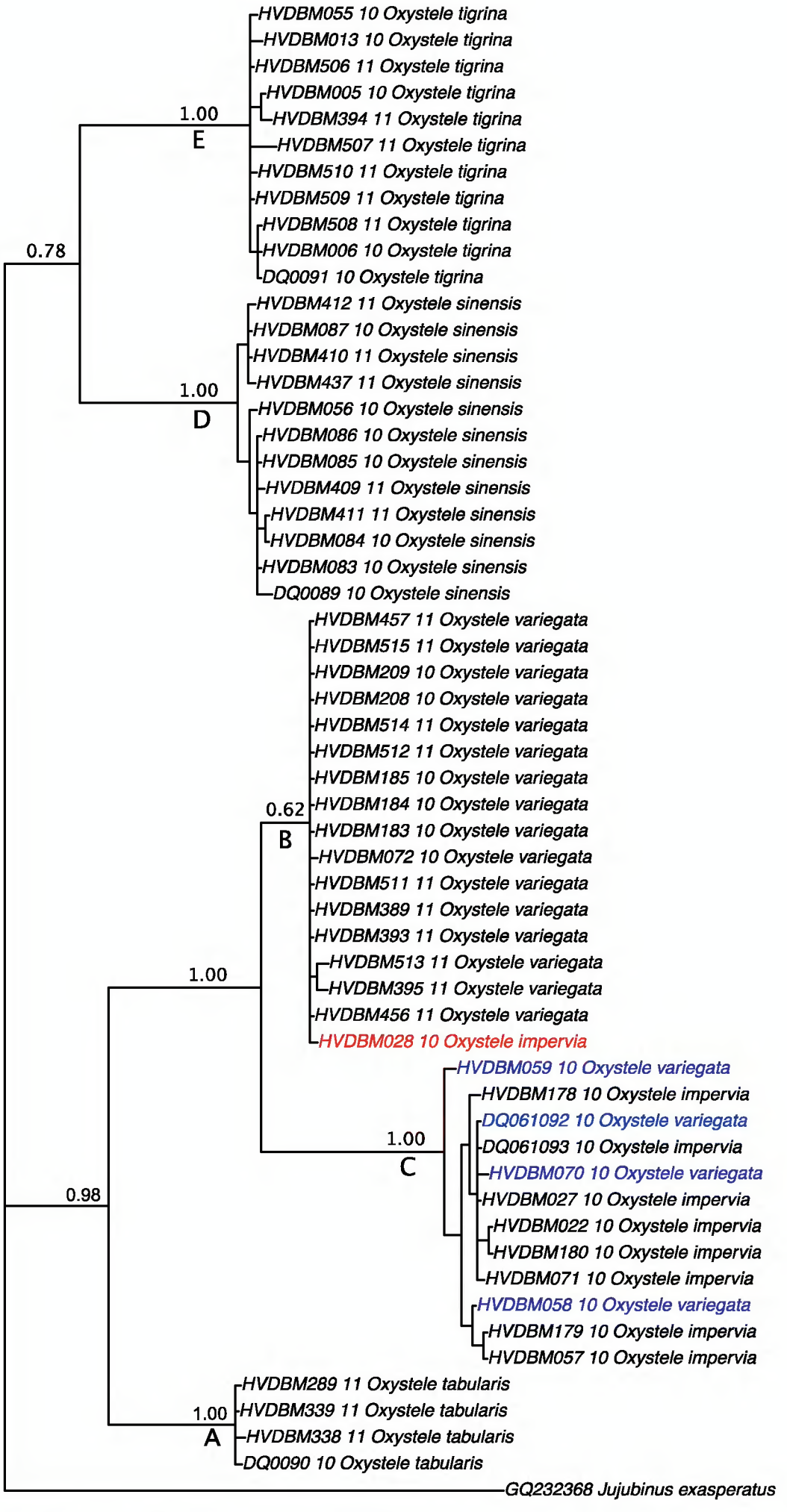

Figure S2. Bayesian tree assembled using MrBayes indicating the groupings of specimens and the posterior probability of the nodes.